УДК 681.3.01

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ В ОЦЕНКЕ КАЧЕСТВА ОБУЧЕНИЯ

О.Г. Берестнева, О.В. Марухина

Томский политехнический университет
E-mail: olgmik@exoft.ru

Рассматривается информационная система оценки качества обучения студентов в вузе. Представлены результаты экспериментальных исследований, проведенных в Томском политехническом университете.

Оценка качества обучения является одним из основных факторов повышения эффективности образовательного процесса в высшей школе в целом. В условиях обеспечения индивидуального подхода к каждому обучаемому очень важно корректно провести качественный и количественный анализ их знаний и умений [1]. Такой анализ является сложной многофакторной зависимостью с большим числом переменных. Проведение подобного анализа часто требует больших затрат сил и времени на проведение статистических расчетов. Оценивание качества обучения с использованием компьютерных технологий позволяет значительно сократить время и трудозатраты анализа и повышает информативность результатов.

Разработанная авторами технология оценки качества обучения студентов базируется на экспертно-статистических алгоритмах, предназначенных для решения прикладных задач, и позволяющих создать единый комплекс программного обеспечения. Результаты анализа могут быть представлены в виде таблиц и графиков.

Программный комплекс имеет дружественный пользователю интерфейс, гибкую систему настроек, сопровождается пакетом справочной информации и представляет собой набор стандартных Windows-приложений, отвечающих всем современным требованиям программного продукта для операционных систем Windows 95/98/NT/2000. Укрупненная схема комплекса представлена на рис. 1.

Все базовые функции системы разделены на два логических блока:

1. Блок компьютерного тестирования.
2. Универсальный блок анализа результатов экспертного оценивания, подробное описание которого приведено в [2, 3].

Блок компьютерного тестирования предназначен для обработки данных педагогического и психологического тестирования и является базовым в *системе поддержки принятия решения по результатам компьютерного тестирования*.

Одной из нерешенных проблем, с которыми в настоящее время в той или иной мере сталкиваются все вузы, является проверка остаточных знаний студентов. Процедура и условия ее проведения, как правило, недостаточно продуманы, а в ряде случаев к ней относятся как к формальному дежурному мероприятию. Главными и типичными недостатками проверки являются подмена остаточных знаний

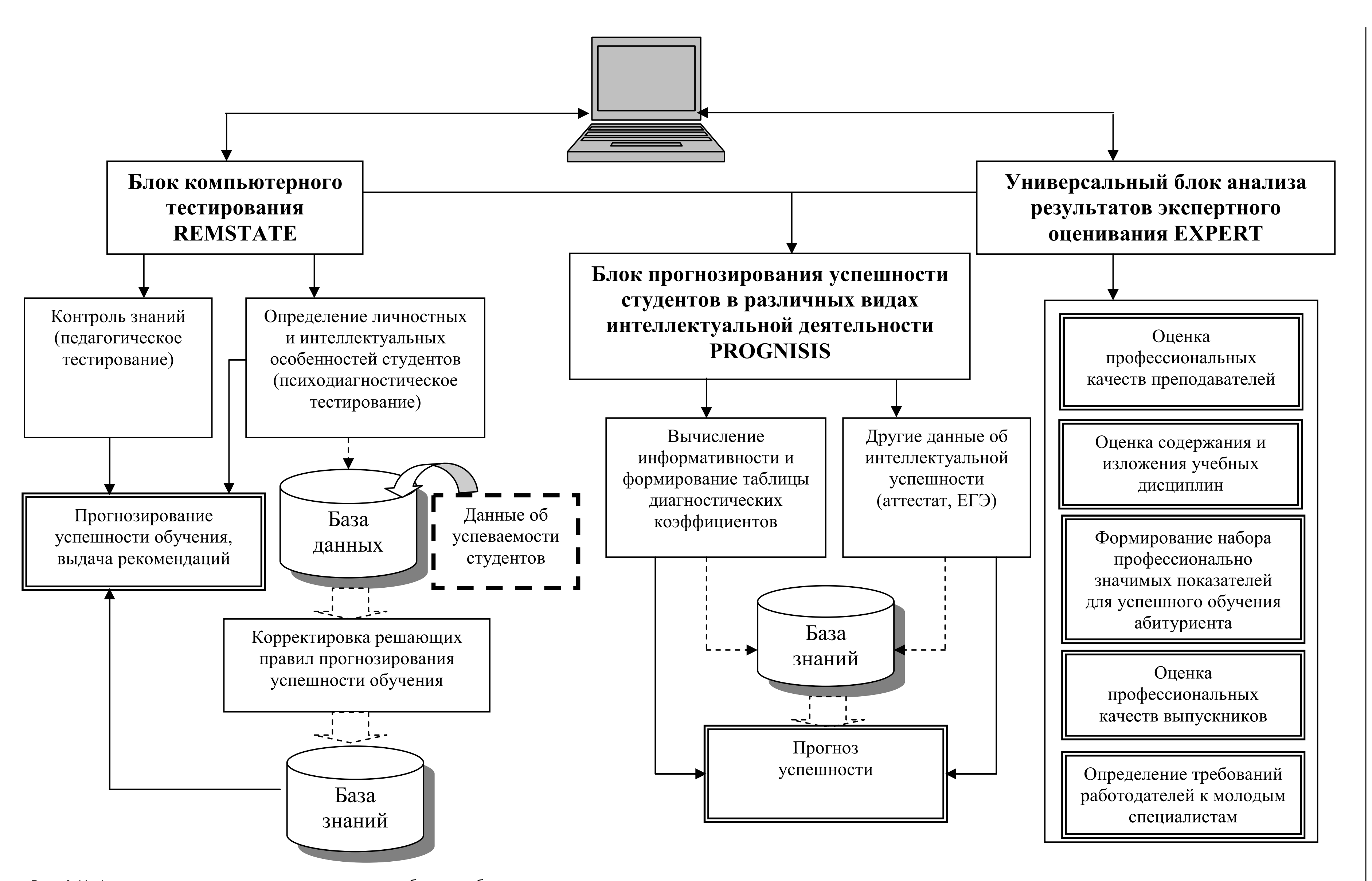


Рис. 1. *Информационная технология оценки качества объектов образовательного процесса*

в качестве ее объекта знаниями "текущими" и списывание студентами ответов на поставленные вопросы. В связи с этим в настоящее время для проверки остаточных знаний все чаще используется метод педагогического тестирования [4].

Тестирование на основе компьютерных технологий официально признано основным средством диагностики качества знаний в проекте Федерального закона "О государственном образовательном стандарте основного общего образования" [5]. Современный уровень развития вычислительной техники и внедрение компьютеров в учебный процесс делают возможным автоматизацию тестового контроля.

В настоящее время существует огромное количество компьютерных систем для контроля знаний студентов [4, 5, 6]. Основным недостатком существующих программных продуктов является отсутствие блока анализа результатов тестирования (или существование его в автономном виде).

На рис. 1 представлена разработанная авторами информационная технология оценки качества обучения. Для оценки остаточных знаний разработана программа REMSTATE, которая включает два блока – генератор тестов (с возможностью конструирования всех существующих форм тестовых заданий) и анализатора, включающего модуль статистической обработки и анализа результатов тестирования.

Оценка и анализ результатов педагогического тестирования

Успех участника тестирования в решении определенного тестового задания зависит, в основном, от двух факторов: *трудности задания* и *подготовленности испытуемого* [6]. При этом только в тривиальной ситуации можно утверждать наверняка, что определенный человек решит определенное задание верно (неверно). В общем случае успех в решении заданий теста имеет вероятностный характер и, следовательно, можно говорить о его количественном измерении некоторым числом p из отрезка [0...1]. Таким образом, вероятность того, что определенный участник тестирования верно решит определенное задание, представляет собой *функцию успеха* [6] двух аргументов – уровня подготовленности испытуемого s и уровня трудности данного задания t

$$P = p(s, t) = p_1(\xi), \text{ где } \xi = s/t.$$

Простейшая модель вероятности успеха, представляющая собой однородную функцию 1-го порядка предложена *датским математиком Г. Рашем* (*G. Rasch*) [7, 8] и имеет вид:

$$p = p(s,t) = \frac{s}{s+t} = \frac{s/t}{1+s/t} = \frac{\xi}{1+\xi}. \quad (1)$$

Функция (1) позволяет наглядно интерпретировать процесс выполнения задания трудности t участником тестирования с уровнем подготовленности s с помощью классической модели теории вероятностей. Если $\ln s = \theta$, $\ln t = \delta \Rightarrow s = e^\theta$, $t = e^\delta$, то функция успеха (1) принимает вид

$$p = \frac{e^\theta}{e^\theta + e^\delta} = \frac{1}{1+e^{\delta-\theta}} \quad (2)$$

и называется основной логистической моделью Раша.

Здесь p – вероятность того, что участник тестирования с уровнем подготовленности θ правильно выполнит задание трудности δ. Вероятность успеха зависит только от одного параметра – разности (θ–δ), и потому модель (2) является однопараметрической [6].

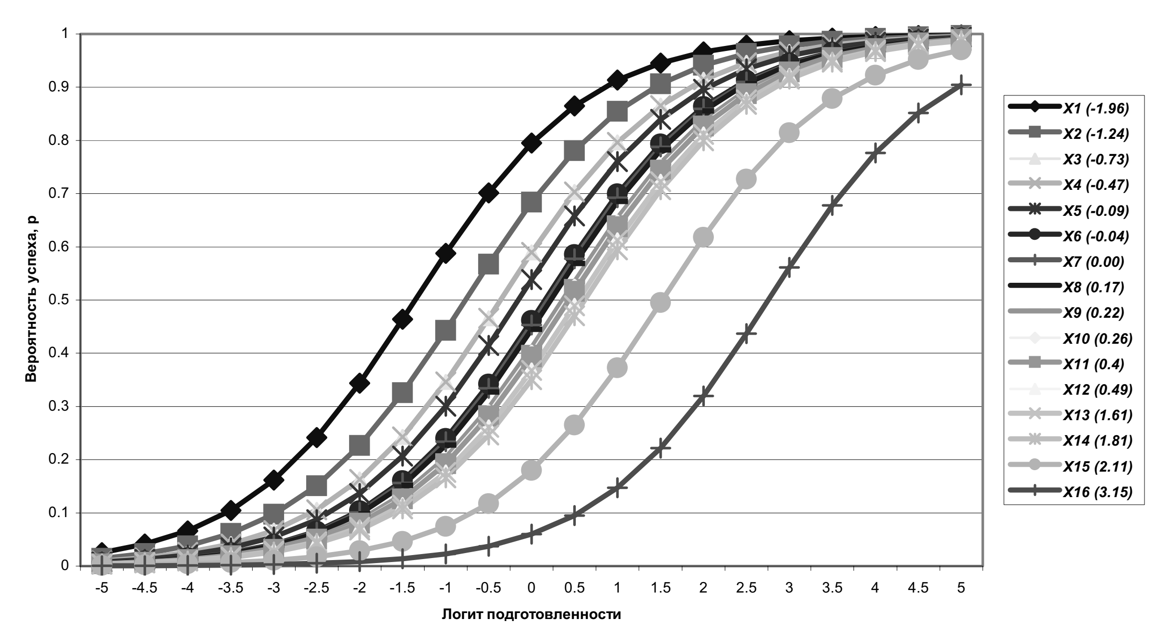


Рис. 2. *Характеристические кривые трудности заданий теста по модели Раша*

На рис. 2 представлен результат обработки данных тестирования по дисциплине "социология" студентов IV курса ТПУ на основе модели Раша [3]. Тест состоял из 16 заданий, которые в дальнейшем тексте будут обозначены как X_i.

Характеристические кривые демонстрируют вероятность правильного ответа на задание опреде-

нем подготовки θ. По построенным графикам определяется трудность задания.

В данном случае наиболее трудным по отношению к остальным оказалось задание, имеющее уровень трудности 3,15, а самым легким – с уровнем трудности – 1,96. На эти задания рекомендуется обратить особое внимание разработчику теста.

Формулу Раша (2) для условной вероятности правильного выполнения j-ого задания теста испытуемыми с различными значениями θ в случае двухпараметрической модели А. Бирнбаума можно переписать в виде [7]

$$p_i = \frac{1}{1+e^{-a_j(\theta_i-\delta_j)}}. \qquad (3)$$

В формуле (3) кроме прежних обозначений вводится новое a_j – для второго параметра j-ого задания теста.

При геометрической интерпретации первый параметр δ_j можно рассматривать как характеристику положения кривой j-го задания относительно оси θ. Второй параметр a_j связан с крутизной кривой задания в точке ее перегиба. А именно значение a_j прямо пропорционально тангенсу угла наклона касательной к характеристической кривой задания теста в точке $\theta = \delta_j$. Это означает, что более крутые кривые соответствуют большим значениям a_j, соответственно для пологих кривых $a_j \to 0$. Таким образом, значения a_j, близкие к нулю, соответствуют случаю, когда испытуемые с разными уровнями подготовки правильно отвечают на j-ое задание с приблизительно равной вероятностью, что, естественно, противоречит ожидаемым прогнозам разработчика теста. Эти задания оказываются бесполезными при дифференциации испытуемых группы по оцениваемому параметру, так как они не несут информации об индивидуальных различиях студентов [4].

Еще более бесполезны задания с отрицательными значениями a_j. На них отвечают правильно с большой вероятностью испытуемые с низким уровнем подготовки, а для знающих студентов с большими значениями θ вероятность правильного ответа стремится к нулю. Число заданий в тесте должно сокращаться в первую очередь за счет устранения таких неудачных заданий даже в том случае, когда другие их характеристики устраивают разработчика теста. Как правило, такое сокращение приводит к повышению надежности и валидности теста.

Проведенный анализ выявляет роль параметра a_j при дифференциации испытуемых. Соответственно параметр a_j получил название *дифференцирующей способности* j-ого задания теста [4]. Формула для оценки параметра дифференцирующей способности j-ого задания теста имеет вид:

$$a_j = \frac{r_{bis_j}}{\sqrt{1-(r_{bis_j})^2}},$$

где r_{bis_j} – бисериальный коэффициент корреляции j-ого задания.

Теоретически значения параметра a_j могут изменяться в интервале (–∞...+∞), но практически

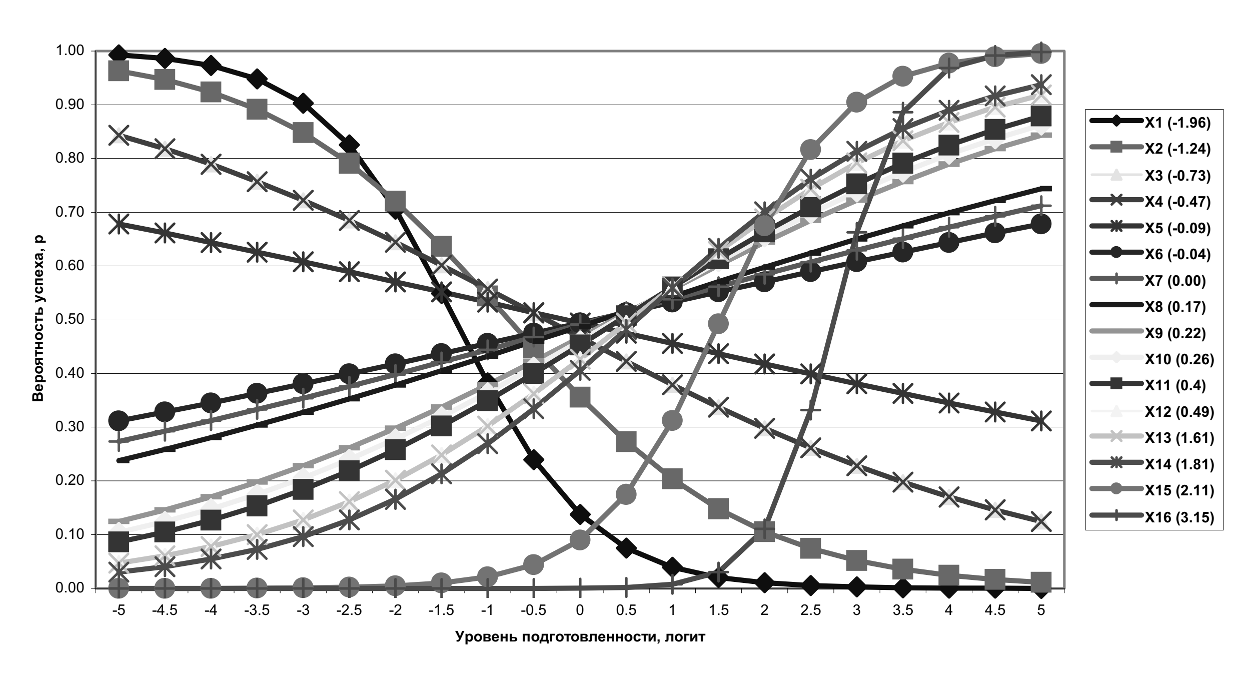


Рис. 3. *Характеристические кривые трудности заданий теста по модели Бирнбаума*

далеко не все эти задания можно включать в тест. Анализ характеристических кривых заданий одинаковой трудности, но разной крутизны позволяет отобрать лучшие задания и определить разумные границы интервала для значений параметра a_j.

Отбор заданий с большими значениями a_j является одним из важных принципов при конструировании нормативно-ориентированного теста. Минимизация длины теста за счет удаления части лишних заданий равной трудности строится на сравнительном анализе крутизны характеристических кривых с точкой перегиба, имеющий одинаковые координаты. Процесс минимизации позволяет выделить одно наиболее эффективное задание с наибольшим значением параметра a_j. На практике рекомендуется, как правило, оставлять задания со значениями a_j, лежащими в интервале (0,5...2,5) [4]. Значение $a_j = 1$ соответствует однопараметрической модели Раша.

На рис. 3 представлены результаты обработки данных тестирования студентов IV курса ТПУ по дисциплине "социология" на основе модели Бирнбаума [3].

В таблице представлены значения параметра a_j, для каждого задания теста X_i.

Таблица. *Значения параметра a_j для каждого задания X_i*

X_{13}	X_{16}	X_7	X_{11}	X_3	X_9	X_4	X_{12}
0,34	0,42	1,07	0,82	0,74	0,93	0,63	1,13
X_2	X_{15}	X_{10}	X_5	X_{14}	X_6	X_1	X_8
-0,14	0,63	0,18	0,65	0,81	0,72	0,24	0,04

Таким образом, применение моделей Раша и Бирнбаума (для анализа данного теста) позволило сделать следующие выводы.

1. Подавляющее большинство заданий теста имеют высокую дифференцирующую способность, однако они неравномерно расположены относительно оси подготовленности θ (что свидетельствует об их однородности относительно показателя трудности и нарушает постулат о том, что тест – система заданий нарастающей сложности). Следовательно, необходимо некоторые задания исключить из теста. Предполагается исключить самые легкие и самые сложные задания, а также задания, имеющие идентичный показатель трудности. Для этого необходимо определить дифференцирующую способность каждого задания.
2. Построение модели Бирнбаума позволяет решить задачу определения дифференцирующей способности заданий теста. В нашем эксперименте одинаковую трудность имеют задания X_{11} и X_7, X_{14} и X_5 – это определяется по значениям логитов трудности. Рекомендуется исключить из теста задания X_{11} (т.к. $a_{X11} < a_{X7}$) и X_5 (т.к. $a_{X5} < a_{X14}$).
3. Рекомендуется исключить задание X_2 ($a_{X2} < 0$), а также задания X_8 ($a_{X8} = 0,04$), X_{10} ($a_{X10} = 0,18$), X_1 ($a_{X1} = 0,24$), X_{13} ($a_{X13} = 0,34$).

Включение в систему тестирования, логистических моделей Раша и Бирнбаума обусловливает важные достоинства, среди которых, прежде всего, необходимо отметить следующие.

1. Превращение измерений, сделанных в дихотомических и порядковых шкалах в линейные измерения. Это позволяет использовать широкий спектр статистических процедур.
2. Оценка трудности тестовых заданий не зависит от выборки испытуемых, на которых была получена; оценка уровня знаний испытуемых не зависит от используемого набора тестовых заданий. Пропуск данных для некоторых комбинаций (испытуемый – тестовое задание) не является критическим.
3. Система тестирования достаточно проста, по сравнению с другими аналогичными системами она характеризуется наименьшим числом параметров.
4. Модели опираются на четкие и конструктивные понятия "трудность задания" и "уровень знаний". Так, одно задание считается более трудным, чем другое, если вероятность правильного ответа на первое задание меньше, чем на второе, независимо от того, кто их выполняет. Аналогично, вероятность правильно ответа более подготовленного студента выше, чем менее подготовленного.
5. Благодаря простой структуре модели существуют удобные вычислительные процедуры для многоаспектной проверки адекватности модели: для всего набора тестовых результатов, для каждого испытуемого, для каждого задания и для каждого конкретного ответа.

Применение результатов психодиагностического тестирования при оценке качества обучения

На рис. 1 показано, что разработанная авторами информационная технология оценки качества обучения включает в себя как педагогическое, так и психодиагностическое тестирование. Остановимся более подробно на особенностях использования компьютерных технологий при анализе результатов психодиагностического тестирования.

Данные психодиагностического тестирования, в основном, измеряются в трех измерительных шкалах: номинальной, ранговой и количественной. Поэтому возникла необходимость разработать алгоритм обработки и анализа результатов психодиагностического тестирования, который бы позволял работать одновременно со всей совокупностью разнородных признаков.

Наличие разнотипной информации в описании состояния объекта при решении задач классификации и диагностики может создать для исследователя дополнительные трудности [7, 8]. Для решения данных задач оценки качества обучения студентов вуза авторами разработана система поддержки принятия решения по результатам компьютерного тестирования, схема которой представлена на рис. 4.



Рис. 4. *Схема анализа результатов тестирования*

Представленная на рис. 4 схема анализа результатов тестирования программно реализована и включена в качестве отдельных модулей (каждый из которых может работать и в автономном режиме) в разработанное авторами программное обеспечение [3].

Для работы с разнотипными признаками разработан алгоритм принятия решений на основе использования неоднородной последовательной процедуры распознавания Е.В. Гублера [3]. Достоинством выбранного нами подхода является возможность перехода от количественной и ранговой шкал к номинальной шкале с произвольным количеством градаций исследуемых признаков, в отличие от известных методов, в которых возможен переход только к дихотомической шкале [3].

На первом этапе проводится корреляционный анализ имеющихся признаков. В зависимости от типа измерительной шкалы использовались: коэффициент корреляции Пирсона, ранговый коэффициент корреляции Спирмена, бисериальный коэффициент корреляции, рангово-бисериальный коэффициент корреляции, коэффициент ассоциации, коэффициент взаимной сопряженности. На втором этапе проводится преобразование исходных признаков в зависимости от измерительной шкалы и формирование матрицы признаков на основе корреляционного анализа и полученных значений информативности каждого признака. На третьем этапе происходит построение матрицы диагностических коэффициентов и вывод решающих правил для принятия решения.

С использованием описанного алгоритма решены задачи прогнозирования [2, 3]:

- успешности обучения студентов младших курсов на основе анализа личностных качеств;
- успешной интеллектуальной самореализации студентов старших курсов на основе анализа показателей познавательной деятельности.

Результаты как психологического, так и педагогического тестирования чаще всего выражены в виде тестовых баллов. Однако, окончательная интерпретация результатов тестирования связана с отнесением полученной числовой характеристики к одной из следующих градаций: очень низкий, низкий, средний, высокий и очень высокий уровень измеряемого качества. При установлении граничных точек для данных интервалов присутствует элемент нечеткости и неопределенности. Таким образом, результат тестирования можно рассматривать как лингвистическую переменную, и использовать для дальнейшего анализа математический аппарат теории нечетких множеств.

Для решения задач оценки качества обучения наиболее эффективным оказался алгоритм принятия решения в условиях неопределенности, основанный на использовании метода анализа альтернатив (принятие решений в условиях неопределенности) [3].

При использовании алгоритма принятия решения в условиях неопределенности оценки объектов задаются как степени соответствия альтернатив понятиям, определяемым критериями. На основе алгоритма принятия решения в условиях неопределенности разработана программа для выбора абитуриентом по результатам компьютерного психодиагностического тестирования наиболее подходящего факультета.

Разработанные нами алгоритмы программно реализованы и внедрены в Томском политехническом университете, а также других томских вузах [3].

Работа частично поддержана грантом РФФИ, проект 03-06-80128

СПИСОК ЛИТЕРАТУРЫ

1. Иванкина Л.И., Берестнева О.Г., Пермяков О.Е. Современный технический университет: Философский и психолого-социологический аспект исследования состояния и развития университета. — Томск: Изд-во Том. ун-та, 2003. — 110 с.
2. Берестнева О.Г., Марухина О.В. Методы многомерного анализа данных в задачах оценки качества образования // Радиоэлектроника. Информатика. Управление. — 2002. — № 1. — С. 15—26.
3. Марухина О.В. Алгоритмы обработки информации в задачах оценивания качества обучения студентов вуза на основе экспертно-статистических методов: Дис. ... канд. техн. наук: 05.13.01. — Томск, 2003. — 165 с.
4. Челышкова М.Б. Теория и практика конструирования педагогических тестов: Учебное пособие. — М.: Логос, 2002. — 432 с.
5. Минин М.Г. Диагностика качества знаний и компьютерные технологии обучения. — Томск: Изд-во ТГПУ, 2000. — 216 с.
6. Нейман Ю.М., Хлебников В.А. Введение в теорию моделирования и параметризации педагогических тестов. — М.: Прометей, 2000. — 168 с.
7. Берестнева О.Г., Иванкина Л.И., Марухина О.В., Пермяков О.Е. Концепция качества образования в техническом вузе // Качество образования: Системы управления, достижения, проблемы: Материалы V Междунар. научно-методич. конф. — Новосибирск: Изд-во НГТУ, 2003. — Том 1. — С. 64—68.
8. Марухина О.В., Берестнева О.Г. Системный подход к оценке качества образования // Стандарты и качество. — 2002. — № 4. — С. 35—36.